

Crystal Structure of Thioesterase SgcE10 Supporting Common Polyene Intermediates in 9- and 10-Membered Eneidyne Core Biosynthesis

Thibault Annaïval,[†] Jeffrey D. Rudolf,[†] Chin-Yuan Chang,[†] Jeremy R. Lohman,[†] Youngchang Kim,^{||,⊥} Lance Bigelow,^{||} Robert Jedrzejczak,^{||} Gyorgy Babnigg,^{||} Andrzej Joachimiak,^{||,⊥,#} George N. Phillips, Jr.,[∇] and Ben Shen^{*,†,‡,§,||}

[†]Department of Chemistry, [‡]Department of Molecular Medicine, and [§]Natural Products Library Initiative at The Scripps Research Institute, The Scripps Research Institute, 130 Scripps Way, Jupiter, Florida 33458, United States

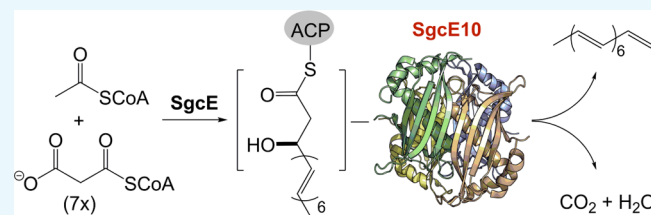
^{||}Midwest Center for Structural Genomics and [⊥]Structural Biology Center, Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, United States

[#]Department of Biochemistry and Molecular Biology, University of Chicago, 929 E. 57th Street, W225, Chicago, Illinois 60637, United States

[∇]BioSciences at Rice and Department of Chemistry, Rice University, 6100 Main Street, Houston, Texas 77251, United States

Supporting Information

ABSTRACT: Eneidyne are potent natural product anticancer antibiotics, and are classified as 9- or 10-membered according to the size of their eneidyne core carbon skeleton. Both 9- and 10-membered eneidyne cores are biosynthesized by the eneidyne polyketide synthase (PKSE), thioesterase (TE), and PKSE-associated enzymes. Although the divergence between 9- and 10-membered eneidyne core biosynthesis remains unclear, it has been observed that nascent polyketide intermediates, tethered to the acyl carrier protein (ACP) domain of PKSE, could be released by TE in the absence of the PKSE-associated enzymes. In this study, we determined the crystal structure of SgcE10, the TE that participates in the biosynthesis of the 9-membered eneidyne C-1027. Structural comparison of SgcE10 with CalE7 and DynE7, two TEs that participate in the biosynthesis of the 10-membered eneidyne calicheamicin and dynemicin, respectively, revealed that they share a common α/β hot-dog fold. The amino acids involved in both substrate binding and catalysis are conserved among SgcE10, CalE7, and DynE7. The volume and the shape of the substrate-binding channel and active site in SgcE10, CalE7, and DynE7 confirm that TEs from both 9- and 10-membered eneidyne biosynthetic machineries bind the linear form of similar ACP-tethered polyene intermediates. Taken together, these findings further support the proposal that the divergence between 9- and 10-membered eneidyne core biosynthesis occurs beyond PKSE and TE catalysis.



INTRODUCTION

Eneidyne natural products are some of the most potent anticancer antibiotics known to date.¹ Eneidyne are classified as 9- or 10-membered according to the size of their eneidyne cores, which are carbocyclic rings consisting of two acetylenic groups conjugated to a double bond or incipient double bond.² The eneidyne cores are biosynthesized by the eneidyne polyketide synthase (PKSE), thioesterase (TE), and PKSE-associated enzymes.³ The iteratively acting PKSE, responsible for initiating eneidyne core biosynthesis, is composed of six domains: ketosynthase (KS), acyltransferase (AT), acyl carrier protein (ACP), ketoreductase (KR), dehydratase (DH), and phosphopantetheinyl transferase (PPTase).⁴ These domains catalyze the decarboxylative condensation of one acetyl-CoA and (minimally) seven malonyl-CoA, as well as associated reductive modifications of the resulting β -ketone groups, to form the ACP-tethered polyketide intermediates for eneidyne

core biosynthesis. Convergent attachment of the various peripheral moieties to the eneidyne core completes eneidyne biosynthesis (Figure 1). Although it is not known if the PKSE-associated enzymes act on the ACP-tethered or discrete forms of the polyketide intermediates to yield the 9- or 10-membered eneidyne core, the ACP-tethered intermediates can be released by the TEs as discrete polyene products in the absence of the PKSE-associated enzymes.⁵ Although the biosynthesis of the peripheral moieties has been extensively studied and is well established,^{6,7} it remains unclear how the eneidyne cores are constructed and what controls the divergence between 9- and 10-membered eneidyne core biosynthesis.

Received: July 5, 2017

Accepted: August 16, 2017

Published: August 30, 2017

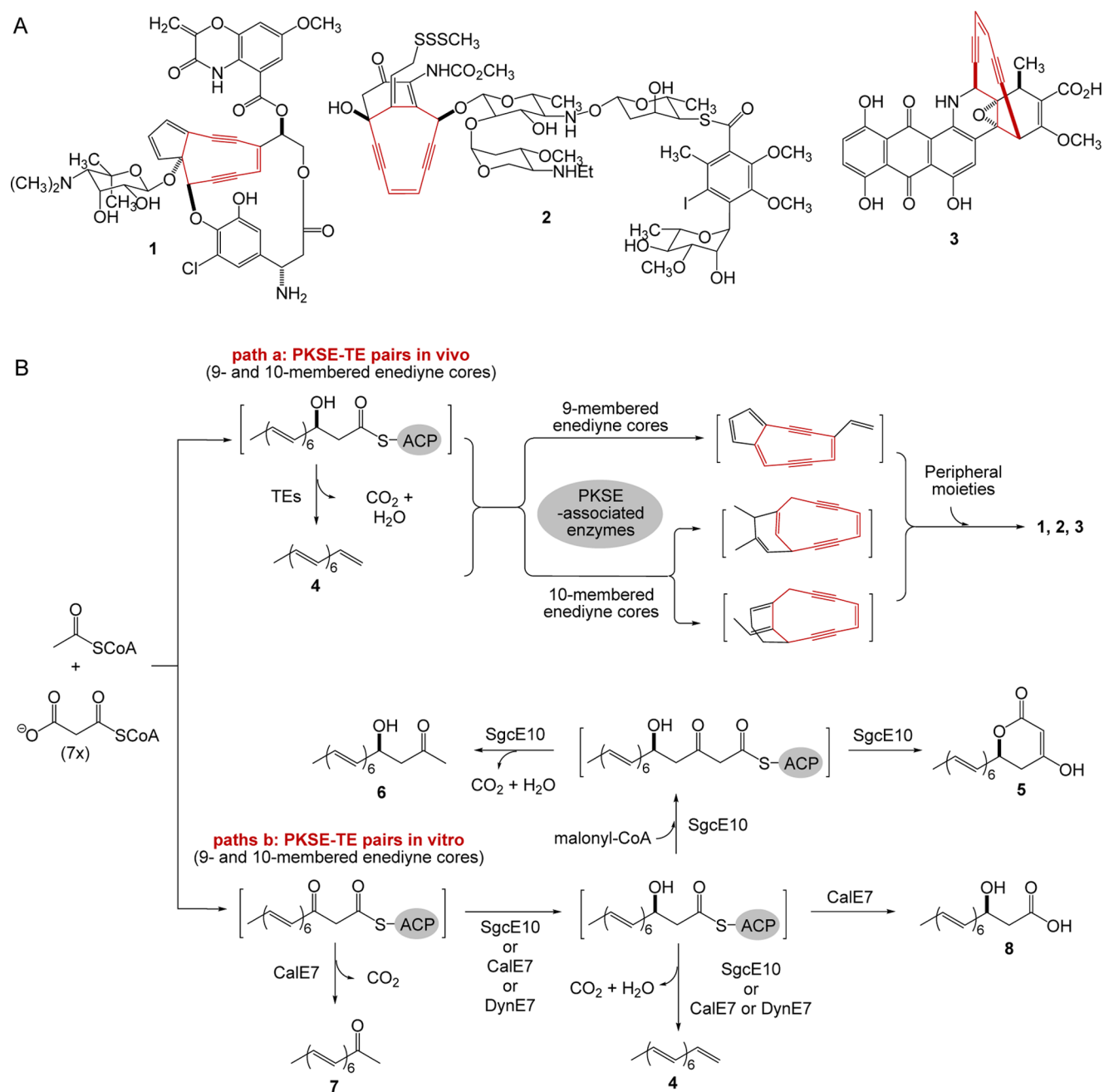


Figure 1. Proposed biosynthesis of enediyne natural products, featuring the PKSE, TE, and PKSE-associated enzymes for both 9- and 10-membered enediyne cores and convergent attachment of the peripheral moieties to the enediyne cores. (A) Structures of selected 9-membered enediyne C-1027 (**1**) and 10-membered enediynes calicheamicin (**2**) and dynemicin (**3**). (B) PKSE–TE pair as a model system for enediyne core biosynthesis. In vivo expression of cognate or mismatched PKSE–TE pairs from selected 9- and 10-membered enediyne biosynthetic machineries all produced **4** as the dominant product (path a). In vitro reconstitution of PKSE–TE pairs from 9- and 10-membered enediyne biosynthetic machineries yielded varying products, as exemplified by SgcE–SgcE10 for the 9-membered enediyne **1** producing **4**, **5**, and **6**, CalE8–CalE7 for the 10-membered enediyne **2** producing **4**, **7**, and **8**, and DynE8–DynE7 for the 10-membered enediyne **3** producing **4** and **7**, respectively (path b).

The exact inventory of the PKSE-associated enzymes and their functions in morphing the nascent polyketide intermediates into the 9- or 10-membered enediyne cores remain elusive.⁸ Efforts to understand enediyne core biosynthesis are currently limited to using the PKSE–TE pair as a model system (Figure 1B).^{9–13} Thus, coexpression of cognate pairs of PKSE–TE from biosynthetic machineries of either 9- or 10-membered enediynes, including the 9-membered enediyne C-1027 (**1**) and 10-membered enediynes calicheamicin (**2**) and dynemicin (**3**), in both heterologous hosts and known enediyne producers all

produced the same polyketide, 1,3,5,7,9,11,13-pentadecaheptaene (**4**), as the dominant product (Figure 1B, path a).^{9,10} PKSEs and TEs from different 9- and 10-membered enediyne biosynthetic machineries were found to be freely interchangeable, affording functional PKSE–TE pairs without any apparent effect on **4** production.¹⁰ These results led to the conclusion that PKSE and TE catalysis does not direct biosynthetic divergence between 9- and 10-membered enediyne cores. In contrast, in vitro reconstitution of PKSE–TE pairs from 9- or 10-membered enediyne biosynthetic machineries yielded

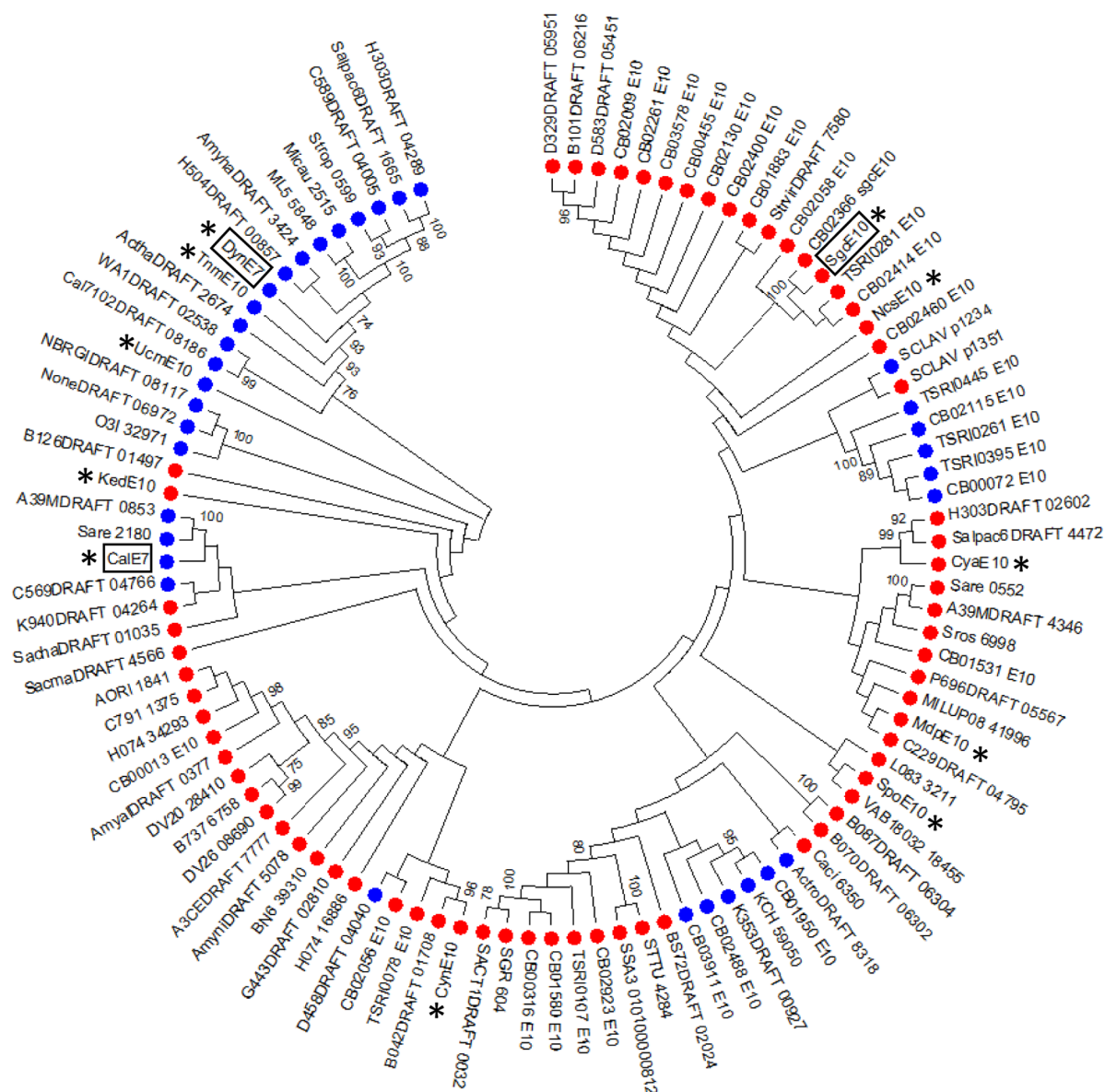


Figure 2. Phylogenetic tree of the 101 TEs from confirmed and predicted 9- and 10-membered enediynes biosynthetic machineries. The TEs from 9-membered enediynes (red) and 10-membered enediynes (blue) biosynthetic machineries show local, but not global, differentiation. Except for the 11 known enediynes (denoted by asterisks), the core size of each enediynes is predicted.²⁰ The three TEs, SgcE10, DynE7, and CalE7 (denoted by black boxes), whose crystal structures have been determined are marked with an asterisk. The sequence alignment was made with ClustalW, and the phylogenetic tree was constructed using the Maximum Likelihood method in MEGA 7.0.18.³⁸

varying products. Although **4** was the common metabolite produced by PKSE–TE pairs from both 9- and 10-membered enediynes biosynthetic machineries, additional products were also characterized, the exact profile of which seemed to be assay-condition dependent.^{11–13} For example, the PKSE–TE pair of SgcE–SgcE10 from the biosynthetic machinery of the 9-membered enediynes **1** produced a nonaketide lactone (**5**) as the main product when SgcE was assayed in the absence or in a low molar ratio of SgcE10. When SgcE was assayed in an excess of SgcE10, **5** and a nonaketide β -hydroxy-ketone (**6**) were observed as the major products, with concomitant accumulation of an increasing amount of **4** (Figure 1B, path b).¹¹ When the PKSE–TE pair of CalE8–CalE7 from the biosynthetic machinery of the 10-membered enediynes **2** was similarly assayed, **4** and the related methyl ketone (**7**) were consistently generated as the major products, with concomitant

accumulation of the corresponding hexaketide β -hydroxy-ketoacid (**8**) as a minor product (Figure 1B, path b).¹² The PKSE–TE pair of DynE8–DynE7 from the biosynthetic machinery of the 10-membered enediynes **3** resulted in the production of **4** and several minor products including **7** and other compounds whose structures were not determined (Figure 1B, path b).¹³ These findings argued that divergence between 9- and 10-membered enediynes core biosynthesis may begin at the stage of PKSE and TE catalysis.¹¹

The crystal structures of CalE7 and DynE7, two TEs from the biosynthetic machineries of the 10-membered enediynes **2** and **3** have been reported.^{14,15} TEs for enediynes core biosynthesis are inherently different from the TEs that are known to catalyze the hydrolytic release of ACP-tethered polyketide intermediates (i.e., type I TEs) or aberrant products (i.e., type II TEs) for reduced polyketide biosynthesis.^{14,15} Type

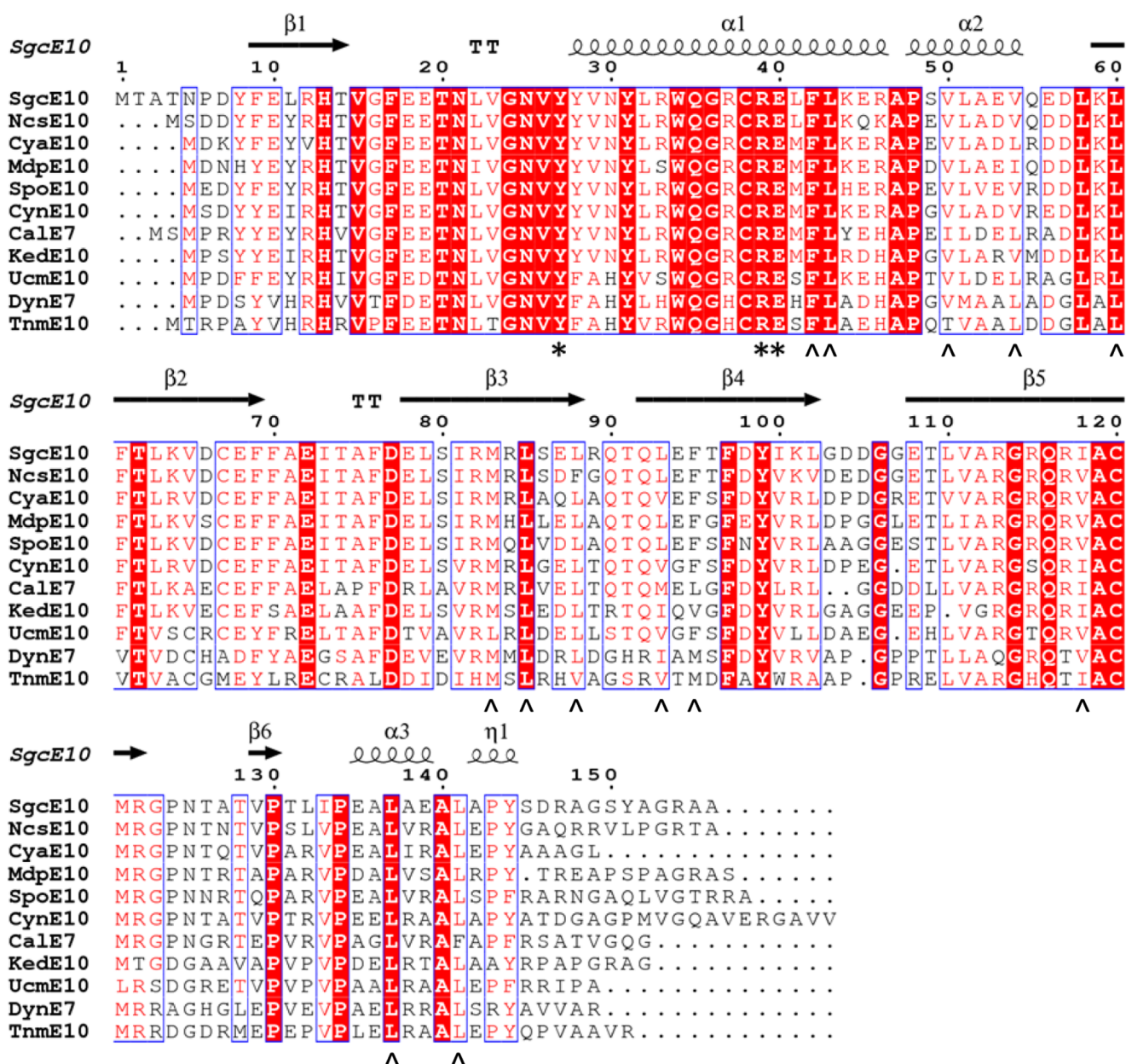


Figure 3. Sequence alignment of the 11 TEs from the seven 9-membered (CyaE10, CynE10, KedE10, MdpE10, NcsE10, SgcE10, and SpoE10) and four 10-membered (CalE7, DynE7, UcmE10, and TnmE10) enediylne biosynthetic machineries. Residues are colored on the basis of the level of conservation (red box with white character shows strict identity, red character similarity, and blue frame similarity across groups). The corresponding secondary structure of SgcE10 is depicted above the sequence alignment. Key residues, involved in substrate binding (denoted by carets) and catalysis (denoted by asterisks) as deduced on the basis of the CalE7 and DynE7 structures,^{14,15} are conserved for all of the TEs.

I TEs are domains of type I polyketide synthases, whereas type II TEs are discrete proteins.^{16,17} Type I and type II TEs share moderate sequence homologies, are monomeric or dimeric, and are members of the well-known α/β hydrolase structural family.^{17–19} In contrast, the TEs for both 9- and 10-membered enediylne core biosynthesis are discrete proteins. They have low sequence homologies with both type I and type II TEs.¹⁵ Both CalE7 and Dyn7 are tetrameric hot-dog fold enzymes.^{14,15} On the basis of their sequence similarity and shared hot-dog folds, the TEs for enediylne core biosynthesis are reminiscent of the acyl-CoA TEs, but with distinct catalytic residues.^{14,15} Whereas the acyl-CoA TEs typically have a catalytic Asp or Glu,^{16,17} the TEs from 9- or 10-membered enediylne biosynthetic machineries possess a conserved Arg, which is implicated in

catalysis;^{14,15} type I and type II TEs for reduced polyketide biosynthesis share a conserved Ser/His/Asp catalytic triad.¹⁸

No structures of TEs from 9-membered enediylne biosynthetic machineries have been reported. Structural comparison between TEs from 9- and 10-membered enediylne biosynthetic machineries could provide a molecular basis to reconcile the differences in product formation observed from in vivo and in vitro studies of selected PKSE–TE pairs for 9- and 10-membered enediylne core biosynthesis. Here, we report the crystal structure of SgcE10, the TE from the 9-membered enediylne biosynthetic machinery of *I* from *Streptomyces globisporus*, revealing that the sequences, structures, oligomerization state, substrate-binding tunnel, active site, and catalytic residues are highly conserved among TEs for 9- (i.e., SgcE10)

and 10-membered (i.e., DynE7 and CalE7) enediyne core biosynthesis. These results further support the proposal that the divergence between 9- and 10-membered enediyne core biosynthesis occurs beyond PKSE and TE catalysis.

RESULTS

Phylogenetic Analysis Revealing Local, but not Global, Differentiation of TEs from 9- and 10-Membered Enediyne Biosynthetic Machineries. Previous bioinformatics analysis of five PKSE–TE pairs (i.e., MdpE–E10, NcsE–E10, and SgcE–E10 for 9-membered enediyne cores and CalE8–E7 and DynE8–E7 for 10-membered enediyne cores) revealed that PKSEs and TEs clade together according to the ring size of their enediyne core.¹⁰ We recently identified 87 enediyne gene clusters, including the gene clusters that encoded the biosynthesis of 10 known enediyne natural products, through a virtual screen of publicly available genome databases.²⁰ We subsequently completed a genome survey of 3400 actinomycetes from The Scripps Research Institute Actinomycetales Collection, discovering 81 additional enediyne producers.²¹ Genome sequencing of 31 representatives of the 81 new enediyne producers confirmed that each harbored a distinct enediyne biosynthetic gene cluster. An enediyne genome neighborhood network was constructed that allowed prediction of 9- or 10-membered enediyne biosynthesis based on the exclusivity of E2 or R3 genes within the enediyne biosynthetic gene clusters. Enediyne discovery was demonstrated by the isolation and structural characterization of tiancimycin, a new 10-membered enediyne natural product.²¹ With the significantly increased number of PKSE–TE pairs now available, we first sought to determine whether the classification of TEs according to their participation in 9- or 10-membered enediyne core biosynthesis still held true.

We constructed a phylogenetic tree using 101 TEs. This included the 11 TEs from the seven 9- and four 10-membered known enediyne biosynthetic machineries and the 60 and 30 TEs from the putative enediyne gene clusters that were predicted to encode the biosynthesis of 9- or 10-membered enediynes, respectively.^{20,21} The TEs range in sequence identity from 38 to 100%. SgcE10 shares 45 and 67% sequence identities with DynE7 and CalE7, respectively, the two TEs from the 10-membered enediyne core biosynthetic machineries whose crystal structures have been determined. DynE7 and CalE7 only share 49% sequence identity, less than that between SgcE10 and CalE7. As with the previous phylogenetic analysis of five TEs,¹⁰ the 11 TEs from the biosynthetic machineries of known enediynes separate into different clades (Figure 2, and also see Figure S1 for similar analysis of PKSEs). In general, the TEs from the 9- or 10-membered enediyne biosynthetic machineries tend to clade together, although most clades are not exclusive. However, there is no clear overall trend where TEs from the 9- or 10-membered enediyne biosynthetic machineries segregate into universal clades (Figure 2).

Despite the differences seen in the phylogenetic analysis, alignment of the TE sequences from the 11 biosynthetic machineries of structurally characterized enediynes revealed sequence conservation of key residues involved in the binding and catalysis of the ACP-tethered polyene intermediates for TEs from both 9- and 10-membered enediyne biosynthetic machineries (Figure 3). On the basis of the structures of CalE7 and DynE7, Phe42, Leu43, Leu60, and Leu137 (numbered based on SgcE10) form the hydrophobic pocket that binds the linear polyene intermediate 4, and Tyr27, Arg39, and Glu40

form a catalytic triad-like motif; these residues are found in all of the 101 TEs.^{14,15} However, no specific residue, conserved only for TEs from the 9- or 10-membered enediyne biosynthetic machineries, is apparent to allow for simple sequence-based classification.

Overall Structure of SgcE10 Establishing a Common Fold of TEs for both 9- and 10-Membered Enediyne Core Biosynthesis. To determine whether the structures of the TEs from the 9- and 10-membered enediyne biosynthetic machineries are different, we next solved the crystal structure of SgcE10, the TE from the 9-membered enediyne biosynthetic machinery of **1** from *S. globisporus*. The best crystals were obtained from the chymotrypsin digested SgcE10 at 16 °C in 0.2 M ammonium tartrate dibasic containing 20% PEG3350. The crystals grew within 1 week and reached sizes of approximately 0.10 mm × 0.02 mm × 0.01 mm. Diffraction analysis revealed that the crystals belong to the $P2_1$ space group with cell dimensions of $a = 57.42$ Å, $b = 184.53$ Å, $c = 59.40$ Å, and $\beta = 95.6^\circ$. The structure was solved by Se-SAD with phasing from three selenomethionines to a resolution of 2.80 Å. The refinement was performed until the structure converged to an R_{work} of 0.211, an R_{free} of 0.270, and a root-mean-square deviation (RMSD) for bond distances and angles of 0.002 Å and 0.604° , respectively. The final model included the protein residues 6–145 of the chains A–E, 6–147 of chain F, 5–146 of chain G, and 5–145 of chain H. There are three tartrates, two glycerols, one ethylene glycol, and 95 ordered water molecules found in the structure. Several residues in both the C- and N-termini were disordered and not modeled.²² Data collection and refinement statistics are summarized in Table 1.

SgcE10 forms a homotetramer with a 222 symmetry (Figure 4A). This agrees well with its tetrameric state in solution, as determined by size-exclusion chromatography (Figure S2). Each monomer has an α/β hot-dog fold that is composed of one β -sheet consisting of five antiparallel strands surrounded by three α -helices (Figure 4B). The four monomers constituting the SgcE10 homotetramer are nearly identical. Structural alignments of monomer A with monomers B–D exhibit RMSD values of 0.3, 0.5, and 0.3 Å, respectively (Figure 4C). Only the N-terminal and C-terminal extremities and the loop at the entrance of the active site are not superimposed for the four monomers, and these three regions have correspondingly high B-factor values (Figure 4D). The SgcE10 homotetramer is a dimer of dimers, with each dimer composed of 10 antiparallel strands surrounding the two central $\alpha 1$ helices (Figure 4A). Each dimer is formed by hydrogen bonds between the two $\beta 2$ strands, the two $\alpha 1$ helices, and the $\alpha 1$ helix and the $\beta 5$ strand, hydrophobic interactions between the $\beta 2$ strand and the $\alpha 1$ helix, and ionic interactions between the two $\beta 2$ strands, and between the $\beta 5$ – $\alpha 3$ loop and the $\beta 1$ – $\alpha 1$ loop. This dimerization of two SgcE10 monomers creates two channels for substrate binding and catalysis (Figure 4E). The SgcE10 tetramer, created by the association of the two dimers, is maintained by hydrophobic interactions and hydrogen bonding between the $\beta 1$ and $\alpha 1$ loops, hydrophobic interactions between the $\alpha 2$ helix and the $\beta 1$ – $\alpha 1$ and $\beta 2$ – $\beta 3$ loops, and ionic interactions between the $\alpha 1$ helix and the $\beta 1$ and $\beta 3$ strands. The association of the two dimers forms a channel with a diameter of ~ 6 Å along the two-fold axis of symmetry (Figure 4A). The tetrameric SgcE10 structure shows RMSD values of 1.3 and 1.0 Å with CalE7 and DynE7, respectively (Figure 4F); structural alignment of DynE7 and CalE7 gives an RMSD of 1.0 Å. Taken together, the structure of SgcE10, in comparison

Table 1. Data Collection and Refinement Statistics

data set	SgcE10
space group	$P2_1$
unit cell (Å)	$a = 57.42, b = 184.53, c = 59.40,$ $\beta = 95.60^\circ$
wavelength (Å)	0.97931
highest resolution bin (Å)	2.85–2.78
number of observed reflections	30 094 (1549) ^a
R_{merge} (%) ^b	12.0 (72.7) ^a
completeness (%)	99.5 (100) ^a
I/σ	10.4 (1.9) ^a
$CC_{1/2}$ (highest resolution shell)	0.659
CC^* ^c	0.891
phasing	
phasing method	SAD
phasing resolution range (Å)	49.8–2.78
number of SeMet	24
overall figure of merit	0.72
refinement	Phenix.refine/refmac5.5
refinement resolution range (Å)	39.16–2.78
R_{work} (%)	21.1
R_{free} (%)	27.0
number of protein residues (model/total)	1125/1272
solvent molecules	95
bond lengths (Å)	0.002
bond angles (deg)	0.604
average B-factors (Å ²)	36.0
Wilson B-factor (Å ²)	69.07
Ramachandran plot (%)	
preferred	95.5
disallowed	0.3
PDB ID	4I4J

^aNumbers in parenthesis are from the highest resolution bin. ^b $R_{\text{meas}} = \sum_{hkl} (N/(N-1))^{1/2} \sum |I_i(hkl) - \bar{I}(hkl)| / \sum_{hkl} \sum I_i(hkl)$. ^c $CC^* = (2CC_{1/2}/(1+CC_{1/2}))^{1/2}$.

with those of CalE7 and DynE7, establishes a common three-dimensional structural fold for TEs from both 9- and 10-membered enediyne biosynthetic machineries.

Active Site of SgcE10 Suggesting Common ACP-Tethered Polyene Intermediates and Their Release Mechanism for Both 9- and 10-Membered Enediyne Core Biosynthesis. The structure of SgcE10 is composed of four active sites, two in each interfacial dimer (Figure 4A,E). The channel formed by dimerization is hydrophobic and is composed of Phe42, Leu43, Val50, Val54, Leu60, Met83, Leu85, Leu88, Leu93, Phe95, Ile118, Leu137, and Leu141. However, the entrance of the channel is open to solvent and is understandably hydrophilic due to the presence of Asn21, Tyr27, Tyr28, Arg39, Glu40, and Thr62 (Figure 5A). We superimposed the A and B chains of the SgcE10 structure with the A and C chains of DynE7 in its apo form (PDB entry 2XFL). The DynE7 residues involved in catalysis (i.e., Tyr23, Arg35, Glu36), as well as those found in the substrate-binding cavity, are present in the SgcE10 structure and similarly positioned (Figure 5B). Only a few minor changes were seen, although the physicochemical properties of these residues were retained (Leu50 to Val54, Ile89 to Leu93, Met91 to Phe95, and Val113 to Ile118 for SgcE10 in comparison with DynE7) (Figure 5B).

Although the CalE7 structure was determined in its apo form,¹⁴ the structures of both the apo and ligand-bound DynE7 forms were obtained, revealing significant conformational changes throughout the substrate-binding tunnel and active site upon complexation with 7.¹⁵ The SgcE10 structure was solved in its apo form (Figures 4 and 5). All attempts to dock 7, 8, or their phosphopantetheinyl thioesters as mimics for the corresponding ACP-tethered polyene intermediates (Figure 1B) into the substrate-binding channel and active site of apo SgcE10 unfortunately yielded no reasonable *in silico* complexes. Considering the conformational changes observed in the structure of DynE7 in complex with 7,¹⁵ similar conformational changes in the apo SgcE10 structure are most likely needed to accommodate the ACP-tethered polyene intermediates. Thus, superposition of the SgcE10 structure and the ligand-bound DynE7 structure, in complex with 7, revealed that the side chain of Phe95 in SgcE10 (Met91 in DynE7) is positioned in the middle of the substrate-binding channel and active site cavity. For proper substrate binding, the side chain of Phe95 must flip to form part of the hydrophobic wall. With no reasonable *in silico* complexes, and in spite of Phe95 blocking part of the substrate-binding channel, overlaying 7 from ligand-bound DynE7 with SgcE10 showed a possible substrate-binding orientation (Figure 5C). The substrate mimic 7 is situated in the SgcE10 substrate-binding channel and active site cavity, interacting with all of the constituent residues via hydrophobic interactions. A water molecule from the SgcE10 structure is positioned beside the ketone group of 7. The oxygen of the latter is pointing toward the side chain of Arg39 at a reasonable distance of 3.6 Å. The ketone oxygen of 7 is equivalent to the β -hydroxyl oxygen of the native substrate of SgcE10 linked to the ACP domain. However, the proximity of the conserved Arg and the water molecule suggests that the position of the ketone oxygen is similar to the thioester oxygen; without an ACP-tethered substrate, it is unclear where the thioester and β -hydroxyl oxygens will be positioned. Arg39 could act as an oxyanion hole, stabilizing the negative charge generated by the attack of the activated water molecule on the thioester bond of the ACP-tethered polyene intermediates. Indeed, Arg39 is positioned via hydrogen bonds with the side chains of Tyr27 and Glu40, which may also be involved in the positioning of the water molecule (3.3 Å) (Figure 5C). The ACP-tethered polyene intermediates are hydrolytically released as either a β -hydroxy- or β -ketoacid, subsequent decarboxylative dehydration or decarboxylation of which finally affords 4 or 7, respectively; it is not known, however, if the latter steps are TE-catalyzed or nonenzymatic (Figure 1B). Generally, TEs and ACPs form protein complexes that facilitate substrate release. However, attempts to dock SgcE10 with the ACP domain of SgcE failed to afford any reasonable complex model, although it has been suggested previously that the TEs bind to the ACP domain-tethered intermediate primarily by recognizing the hydrophobic linear polyene.¹⁵

DISCUSSION

The enediynes are some of the most potent antitumor antibiotics, and their fascinating molecular architectures have inspired numerous biosynthetic studies. Since the cloning of the first 9-membered (1)²³ and 10-membered (2)²⁴ enediyne biosynthetic gene clusters in 2002, a total of 12 enediyne biosynthetic gene clusters, encoding the biosynthesis of seven 9-membered enediynes and five 10-membered enediynes, have now been cloned, sequenced, and partially characterized. Over

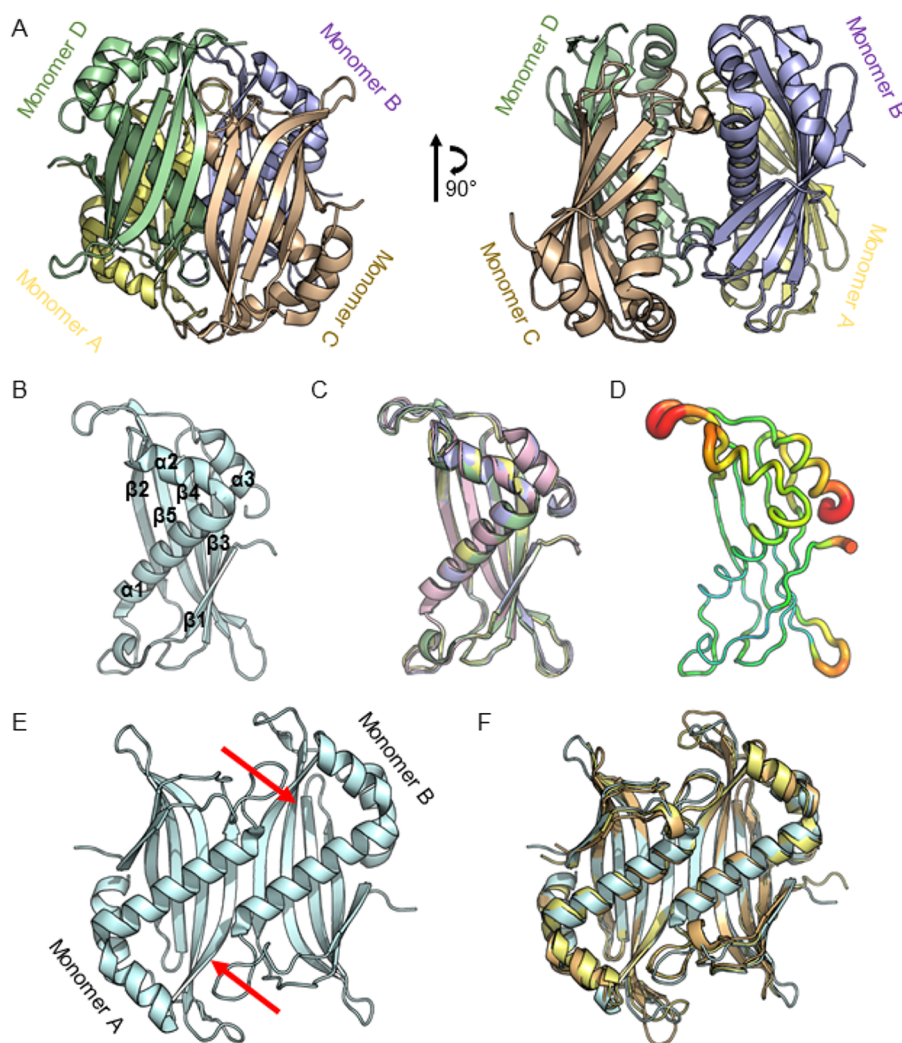


Figure 4. SgcE10 is a homotetramer that shares the same structural fold as that of CalE7 and DynE7. (A) The overall structure of SgcE10 is a homotetramer consisting of a dimer of dimers. (B) The α/β hot-dog monomer fold composed of a β -sheet comprising five antiparallel strands ($\beta 1$ – $\beta 5$) wrapped by three α -helices ($\alpha 1$ – $\alpha 3$). (C) Superposition of the chains A (green), B (blue), C (pink), and D (yellow) comprising the homotetramer of SgcE10 showing that the monomers are identically related to each other. (D) Some weak variations observed in the N- and C-terminal extremities and the loop positioned at the entrance of the active site as depicted with B-factor presentation of the chain A of SgcE10. The warm colors in the extremity and loop regions indicate a high degree of mobility. (E) Dimerization of two monomers affording the dimer with two active sites (red arrows). (F) Superposition of the dimers of SgcE10 (blue), DynE7 (orange), and CalE7 (yellow) establishing a common structural fold for TEs from both 9- and 10-membered enediynes biosynthetic machineries.

100 additional enediyne biosynthetic gene clusters have been discovered through genome mining; although most of these encoded enediyne natural products are yet to be structurally characterized, these discoveries underline the remarkable potential of nature in biosynthesizing the enediyne family of natural products.^{20,21} Tremendous progress has been made in understanding the biosynthesis of the peripheral moieties and uncovering the convergent strategy of their attachment to the enediyne cores to complete enediyne biosynthesis;^{6,7} however, how the enediyne cores are constructed and what controls the divergence between 9- and 10-membered enediyne core biosynthesis remain elusive.

Enediyne core biosynthesis is currently limited to using the PKSE–TE pair as a model system, and all attempts to produce and structurally characterize biosynthetic intermediates, leading to either 9- or 10-membered enediyne cores, have met with little success.^{9–13} Although *in vivo* coexpression of cognate or mismatched PKSE–TE pairs from both 9- and 10-membered

enediyne biosynthetic machineries all resulted in the exclusive production of the same polyene metabolite **4** (Figure 1B, path a), *in vitro* reconstitution of PKSE–TE pairs from selected 9- or 10-membered enediyne biosynthetic machineries yielded varying metabolites, for example, **4**, **5**, and **6** by SgcE–SgcE10 (Figure 1B, path b), **4**, **7**, and **8** by CalE8–CalE7 (Figure 1B, path c), and **4** and **7** by DynE8–DynE7. The metabolite profile discrepancy between the *in vivo* and *in vitro* studies of the PKSE–TE pairs raised the question of whether the varying metabolites observed *in vitro* are relevant to enediyne core biosynthesis and whether TEs play a gate-keeping role, selectively channeling the nascent ACP-tethered polyene intermediates to 9- or 10-membered enediyne core biosynthesis. Because the structures of CalE7 and DynE7, two TEs from the biosynthetic machineries of the 10-membered enediynes **2** and **3**, are known, we reasoned that structural comparison of CalE7 and DynE7 with TEs from 9-membered enediyne biosynthetic machineries may provide insight into the

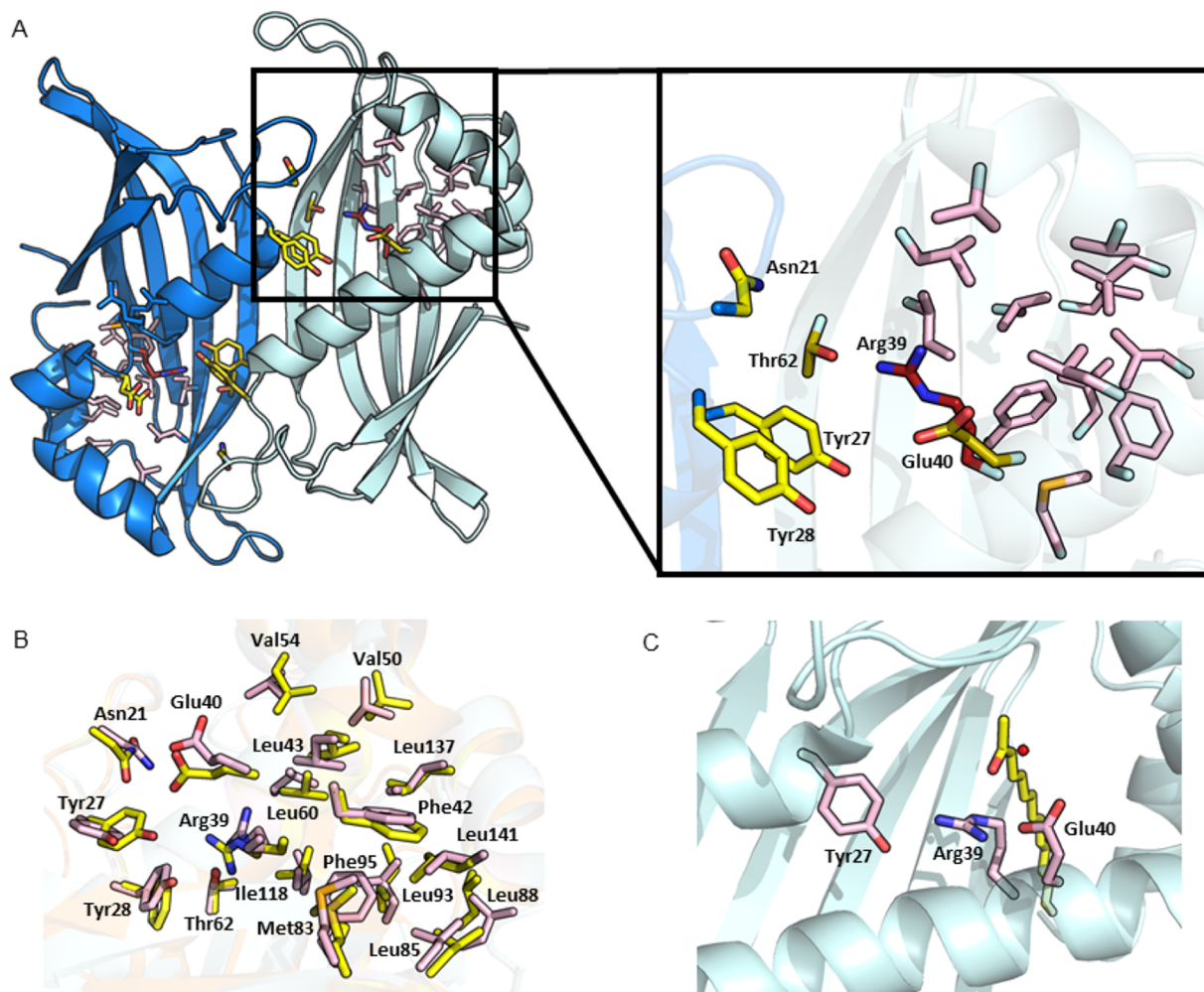


Figure 5. SgcE10 sharing the same substrate-binding tunnel and active site as those of CalE7 and DynE7. (A) The active site of SgcE10 is composed of a hydrophobic pocket (pink) with hydrophilic residues (yellow) with the catalytic Arg (red) at the entrance of the pocket. The hydrophobic pocket is involved in binding of the linear polyene intermediate, while the hydrophilic residues likely accept the phosphopantetheine arm of ACP. The substrate-binding channel is formed by the dimerization of two monomers (light and dark blue); Arg39, Glu40, and Thr62 are from one monomer and Asn21, Tyr27, and Tyr28 are from the other. (B) Superposition of the SgcE10 (blue) and DynE7 (orange) active sites revealing the same substrate-binding and catalytic residues of SgcE10 (pink) as those of DynE7 (yellow). (C) Superposition of 7 from the ligand-bound DynE7 structure (PDB entry 2XEM) with one of the four active sites of SgcE10, supporting a similar mechanism for the TE-catalyzed hydrolysis of the ACP-tethered polyene intermediates in 9- and 10-membered enediyne core biosynthesis. A catalytic water molecule (red sphere) is positioned by Glu40 in front of the ketone and could be involved in the release of the heptane. The negative charge of the tetrahedral intermediate, resulting from nucleophilic attack by the water molecule, could be stabilized by Arg39, which is oriented by Tyr27.

nature of the ACP-tethered polyene intermediates and the roles of the TEs in enediyne core biosynthesis.

Taking advantage of the rapidly growing family of enediyne biosynthetic gene clusters,^{20,21} we first carried out extensive bioinformatics analysis of the PKSE–TE pairs from all 9- and 10-membered enediyne biosynthetic machineries, and this revealed that TEs are generally categorized into clades based on their involvement in 9- or 10-membered enediyne biosynthesis, but that this trend is not strictly followed (Figure 2). However, amino acid sequence alignments showed that all TEs are fairly conserved (38–100% identities) and, most importantly, the residues involved in substrate binding and catalysis are highly conserved (Figure 3), suggesting similar ACP-tethered polyene intermediates and mechanisms of catalysis for TEs in both 9- and 10-membered enediyne core biosynthesis. We next solved the crystal structure of SgcE10 from the 9-membered biosynthetic machinery of 1 and compared it with those of DynE7 and CalE7, the two TEs from the 10-membered

enediayne biosynthetic machineries of 2 and 3. The overall structure (Figure 4) and the key residues lining up the substrate-binding tunnel and active sites are all highly conserved among SgcE10, DynE7, and CalE7 (Figure 5), suggesting that the catalytic processes of these enzymes are very similar. Considering that the PKSE–TE pairs from both 9- and 10-membered enediayne biosynthetic machineries produced the same polyene metabolite 4,¹⁰ which was also detected from the fermentations of all known enediayne producers examined,²⁵ and combined with the new findings that SgcE10 features a nearly identical volume and shape of substrate-binding tunnel and active site cavity to CalE7 and DynE7, it is likely that the ACP-tethered polyene substrates for the TEs and their catalytic release for both 9- and 10-membered enediayne core biosynthesis are identical. Therefore, our current sequence and structural analysis of SgcE10 and other TEs, along with previous *in vivo* and *in vitro* characterizations of PKSE–TE pairs, further support the proposal that the divergence between

9- and 10-membered enediynes core biosynthesis occurs beyond PKSE and TE catalysis. Biosynthetic investigations are currently aimed at characterizing the role of the PKSE-associated enzymes in 9- or 10-membered enediynes core biosynthesis (Figure 1B, path a).

MATERIALS AND METHODS

Gene Cloning. Gene cloning was performed according to previously published procedures.²⁶ Full length *sgcE10* (gi: 24575121; locus version: AAL06692.1) was amplified from the genomic DNA of *S. globisporus* with KOD Hot Start DNA polymerase using 5'-TACTTCCAATCCAATGCCATGACCGCGACGAATCCTGACTA-3' and 5'-TTATC-CACTTCCAATGTTAGGCGGCGCGTCCCCG-3' as forward and reverse primers, respectively. The amplification buffer was supplemented with betaine to a final concentration of 2.5 M. The PCR product was purified and treated with T4 DNA polymerase in the presence of dCTP according to the specifications of the manufacturer (New England Biolabs), cloned into pMCSG57²⁷ according to ligation-independent procedures,²⁸ and transformed into *E. coli* BL21(DE3)-Gold (Stratagene).

Gene Expression and Protein Purification. To produce the SgcE10 protein, a bacterial culture was grown at 37 °C, 190 rpm in 2 L of enriched M9 medium²⁹ until it reached OD₆₀₀ = 1.0. After air-cooling it down to 4 °C for 60 min, inhibitory amino acids (25 mg/L each of L-valine, L-isoleucine, L-leucine, L-lysine, L-threonine, and L-phenylalanine) and 90 mg/L selenomethionine (SeMet) were added. Gene expression was induced with 0.5 mM isopropyl- β -D-thiogalactoside (IPTG). The cells were incubated overnight at 18 °C, harvested by centrifugation, and resuspended in lysis buffer [50 mM HEPES, pH 8.0, containing 5% (v/v) glycerol, 500 mM NaCl, 20 mM imidazole, and 10 mM β -mercaptoethanol]. Cells were disrupted by lysozyme treatment (1 mg/mL) and sonication, and the insoluble cellular material was removed by centrifugation. The SeMet-labeled SgcE10 protein was purified using Ni-NTA affinity chromatography and the ÄKTExpress system (GE Healthcare Life Sciences) with the addition of 10 mM β -mercaptoethanol in all buffers. The N-terminal His₆-tag of SgcE10 was cleaved using recombinant His₆-tagged TEV protease. An additional step of Ni-NTA affinity chromatography was performed to remove the protease, uncut His₆-tagged SgcE10 protein, and the His₆-tag. The oligomeric state of SgcE10 was defined by using size-exclusion chromatography (Superdex 200 16/600 column, GE Healthcare Life Sciences). Pure SgcE10 was concentrated using Amicon Ultra-15 concentrators (Millipore) in 20 mM HEPES buffer, pH 8.0, containing 250 mM NaCl and 2 mM dithiothreitol (DTT). Protein concentrations were determined from the absorbance at 280 nm using a molar absorption coefficient (ϵ_{280} = 15 930 M⁻¹ cm⁻¹).³⁰ The concentration of SgcE10 samples used for crystallization was 42 mg/mL. A second batch of SeMet-labeled SgcE10 was purified for optimization with reductive alkylation and partial proteolysis.²⁶ The concentration of SgcE10 samples used for crystallization was 5 mg/mL for reductively methylated, 29 mg/mL for reductively ethylated, 26 mg/mL for reductively *iso*-propylated, and 48 mg/mL for partially proteolyzed (with chymotrypsin, trypsin, or thermolysin), respectively. Individual aliquots of the purified SgcE10 were stored at -80 °C until needed.

SgcE10 Crystallization. SgcE10 was screened for crystallization conditions with the help of a Mosquito liquid dispenser

(TTP Labtech) using the sitting-drop vapor-diffusion technique in 96-well CrystalQuick plates (Greiner Bio-One). For each condition, 0.4 μ L of SgcE10 and 0.4 μ L of crystallization formulation were mixed, and the mixture was equilibrated against 140 μ L of the reservoir in the well. Several commercially available crystallization screens were used including: MCSG-1-3 (Anatrace) at 24 and 4 °C for the unmodified protein, MCSG-1-4 at 24 °C for the reductively alkylated proteins, and PEG/Ion HT (Hampton Research Corp.) at 16 °C for the partially proteolyzed proteins.

Data Collection and Structure Determination and Refinement. Diffraction data were collected at 100 K at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory.³¹ The single-wavelength data at 0.97931 up to 2.8 Å were collected from a single protein crystal of SgcE10. Integration and scaling were performed with the HKL-3000 program package.³² The structure was determined by single-wavelength anomalous diffraction (SAD) phasing utilizing the anomalous signal from Se atoms with PHENIX (AutoSol and AutoBuild)³³ using the peak data to 2.8 Å. An extensive manual model was built with Coot³⁴ with subsequent refinement using phenix.refine,³³ and Refmac5.5 from the CCP4 suite³⁵ was used. After multiple rounds of refinement with coot and phenix.refine, the structure converged. The stereochemistry of the structure was checked with PROCHECK³⁶ and a Ramachandran plot.³⁷ The atomic coordinates and experimental structure factors of SgcE10 were deposited in the Protein Data Bank (PDB) under the code 4I4J.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.7b00933.

The phylogenetic tree of the 102 PKSEs (including EspE from the biosynthetic machinery of esperamicin) of which 12 were from biosynthetic machineries of the known enediynes natural products, including seven 9- and five 10-membered enediynes, and 60 and 30 were from putative enediynes gene clusters that were predicted to encode the biosynthesis of 9- or 10-membered enediynes, respectively,²⁰ complementing Figure 2 for the 101 TEs (Figure S1); determination of SgcE10 as a homotetramer in solution by size-exclusion chromatography (Figure S2) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: shenb@scripps.edu. Tel: (561) 228-2456. Fax: (561) 228-2472.

ORCID

Ben Shen: 0000-0002-9750-5982

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This is manuscript #29514 from The Scripps Research Institute. This work is supported in part by the National Institute of Health Grants GM094585 (to A.J.) and GM098248 (to G.N.P.), and CA078747 (to B.S.) and GM115575 (to B.S.). The use of Structural Biology Center beamlines at the

Advanced Photon Source was supported in part by the U.S. Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH113 (to A.J.). J.D.R. and C.-Y.C. are supported in part by an Arnold O. Beckman Postdoctoral Fellowship and the Fellowship of Academia Sinica-The Scripps Research Institute Postdoctoral Talent Development Program, respectively.

REFERENCES

- (1) Wang, X.-W.; Xie, H. C-1027. Antineoplastic antibiotic. *Drugs Future* **1999**, *24*, 847–852.
- (2) Shen, B.; Liu, W.; Nonaka, K. Eneidiyne natural products: Biosynthesis and prospect towards engineering novel antitumor agents. *Curr. Med. Chem.* **2003**, *10*, 2317–2325.
- (3) Van Lanen, S. G.; Oh, T. J.; Liu, W.; Wendt-Pienkowski, E.; Shen, B. Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. *J. Am. Chem. Soc.* **2007**, *129*, 13082–13094.
- (4) Nicolaou, K. C.; Dai, W.-M. Chemistry and biology of the enediyne anticancer antibiotics. *Angew. Chem., Int. Ed.* **1991**, *30*, 1387–1416.
- (5) Zhang, J.; Van Lanen, S. G.; Ju, J.; Liu, W.; Dorrestein, P. C.; Li, W.; Kelleher, N. L.; Shen, B. A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1460–1465.
- (6) Liang, Z. X. Complexity and simplicity in the biosynthesis of enediyne natural products. *Nat. Prod. Rep.* **2010**, *27*, 499–528.
- (7) Van Lanen, S. G.; Shen, B. Biosynthesis of enediyne antitumor antibiotics. *Curr. Top. Med. Chem.* **2008**, *8*, 448–459.
- (8) Chen, X.; Ji, R.; Jiang, X.; Yang, R.; Liu, F.; Xin, Y. Iterative type I polyketide synthases involved in enediyne natural product biosynthesis. *IUBMB Life* **2014**, *66*, 587–595.
- (9) Kong, R.; Goh, L. P.; Liew, C. W.; Ho, Q. S.; Murugan, E.; Li, B.; Tang, K.; Liang, Z. X. Characterization of a carbonyl-conjugated polyene precursor in 10-membered enediyne biosynthesis. *J. Am. Chem. Soc.* **2008**, *130*, 8142–8143.
- (10) Horsman, G. P.; Chen, Y.; Thorson, J. S.; Shen, B. Polyketide synthase chemistry does not direct biosynthetic divergence between 9- and 10-membered enediynes. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11331–11335.
- (11) Chen, X.; Guo, Z. F.; Lai, P. M.; Sze, K. H.; Guo, Z. Identification of a nonaketide product for the iterative polyketide synthase in biosynthesis of the nine-membered enediyne C-1027. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 7926–7928.
- (12) Belecki, K.; Crawford, J. M.; Townsend, C. A. Production of octaketide polyenes by the calicheamicin polyketide synthase CalE8: Implications for the biosynthesis of enediyne core structures. *J. Am. Chem. Soc.* **2009**, *131*, 12564–12566.
- (13) Sun, H.; Kong, R.; Zhu, D.; Lu, M.; Ji, Q.; Liew, C. W.; Lescar, J.; Zhong, G.; Liang, Z. X. Products of the iterative polyketide synthases in 9- and 10-membered enediyne biosynthesis. *Chem. Commun. (Camb)* **2009**, 7399–7401.
- (14) Kotaka, M.; Kong, R.; Qureshi, I.; Ho, Q. S.; Sun, H.; Liew, C. W.; Goh, L. P.; Cheung, P.; Mu, Y.; Lescar, J.; Liang, Z. X. Structure and catalytic mechanism of the thioesterase CalE7 in enediyne biosynthesis. *J. Biol. Chem.* **2009**, *284*, 15739–15749.
- (15) Liew, C. W.; Sharff, A.; Kotaka, M.; Kong, R.; Sun, H.; Qureshi, I.; Bricogne, G.; Liang, Z. X.; Lescar, J. Induced-fit upon ligand binding revealed by crystal structures of the hot-dog fold thioesterase in dynemicin biosynthesis. *J. Mol. Biol.* **2010**, *404*, 291–306.
- (16) Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **1999**, *6*, 117–125.
- (17) Heathcote, M. L.; Staunton, J.; Leadlay, P. F. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **2001**, *8*, 207–220.
- (18) Tsai, S. C.; Miercke, L. J.; Krucinski, J.; Gokhale, R.; Chen, J. C.; Foster, P. G.; Cane, D. E.; Khosla, C.; Stroud, R. M. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14808–14813.
- (19) Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **2012**, *29*, 1050–1073.
- (20) Rudolf, J. D.; Yan, X.; Shen, B. Genome neighborhood network reveals insights into enediyne biosynthesis and facilitates prediction and prioritization for discovery. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 261–276.
- (21) Yan, X.; Ge, H.; Huang, T.; Hindra; Yang, D.; Teng, Q.; Crnovčić, I.; Li, X.; Rudolf, J. D.; Lohman, J. R.; Gansemans, Y.; Zhu, X.; Huang, Y.; Zhao, L. X.; Jiang, Y.; Van Nieuwerburgh, F.; Rader, C.; Duan, Y.; Shen, B. Strain Prioritization and Genome Mining for Eneidiyne Natural Products. *MBio* **2016**, *7*, No. e02104-16.
- (22) Kim, Y.; Dementieva, I.; Zhou, M.; Wu, R.; Lezondra, L.; Quartey, P.; Joachimiak, G.; Korolev, O.; Li, H.; Joachimiak, A. Automation of protein purification for structural genomics. *J. Struct. Funct. Genomics* **2004**, *5*, 111–118.
- (23) Liu, W.; Christenson, S. D.; Standage, S.; Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **2002**, *297*, 1170–1173.
- (24) Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; Farnet, C. M.; Thorson, J. S. The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **2002**, *297*, 1173–1176.
- (25) Shen, B.; Hindra; Yan, X.; Huang, T.; Ge, H.; Yang, D.; Teng, Q.; Rudolf, J. D.; Lohman, J. R. Eneidiynes: Exploration of microbial genomics to discover new anticancer drug leads. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 9–15.
- (26) Kim, Y.; Babnigg, G.; Jedrzejczak, R.; Eschenfeldt, W. H.; Li, H.; Maltseva, N.; Hatzos-Skintges, C.; Gu, M.; Makowska-Grzyska, M.; Wu, R.; An, H.; Chhor, G.; Joachimiak, A. High-throughput protein purification and quality assessment for crystallization. *Methods* **2011**, *55*, 12–28.
- (27) Eschenfeldt, W. H.; Stols, L.; Millard, C. S.; Joachimiak, A.; Donnelly, M. I. A family of LIC vectors for high-throughput cloning and purification of proteins. *Methods Mol. Biol.* **2009**, *498*, 105–115.
- (28) Aslanidis, C.; De Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **1990**, *18*, 6069–6074.
- (29) Donnelly, M. I.; Zhou, M.; Millard, C. S.; Clancy, S.; Stols, L.; Eschenfeldt, W. H.; Collart, F. R.; Joachimiak, A. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expression Purif.* **2006**, *47*, 446–454.
- (30) Gill, S. C.; von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989**, *182*, 319–326.
- (31) Rosenbaum, G.; Alkire, R. W.; Evans, G.; Rotella, F. J.; Lazarski, K.; Zhang, R. G.; Ginell, S. L.; Duke, N.; Naday, I.; Lazarz, J.; Molitsky, M. J.; Keefe, L.; Gonczy, J.; Rock, L.; Sanishvili, R.; Walsh, M. A.; Westbrook, E.; Joachimiak, A. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J. Synchrotron Radiat.* **2006**, *13*, 30–45.
- (32) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. HKL-3000: the integration of data reduction and structure solution— from diffraction images to an initial model in minutes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 859–866.
- (33) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213–221.

(34) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

(35) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355–367.

(36) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

(37) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12–21.

(38) Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874.